HIPE – An Energy-Status-Data Set from Industrial Production

Simon Bischof†
simon.bischof2@kit.edu

Holger Trittenbach‡
holger.trittenbach@kit.edu

Michael Vollmer‡
michael.vollmer@kit.edu

Dominik Werle‡
dominik.werle@kit.edu

Thomas Blank†
thomas.blank@kit.edu

Klemens Böhm‡
klemens.boehm@kit.edu

†Karlsruhe Institute of Technology (KIT), Germany
Institute for Data Processing and Electronics (IPE)

‡Karlsruhe Institute of Technology (KIT), Germany
Institute for Program Structures and Data Organization (IPD)

ABSTRACT

Energy-related data sets from industrial production are rare, and related research questions have only attracted little attention. To facilitate research on these challenges, we describe a comprehensive machine-level energy-data set from a production site and make it publicly available. We then sketch applications where our data set may serve as a benchmark and catalyst for further research.

KEYWORDS

Energy Status Data, Smart Meter Data

ACM Reference Format:
Simon Bischof, Holger Trittenbach, Michael Vollmer, Dominik Werle, Thomas Blank, and Klemens Böhm. 2018. HIPE – An Energy-Status-Data Set from Industrial Production. In *Proceedings of e-Energy '18*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3208903.3210278

1 INTRODUCTION

The publication of energy-related data has ignited insightful research on future energy systems. Examples are the disaggregation of consumption data, in order to recommend behavior that saves energy [5, 13], and demand-side management for the integration of renewable energy sources [3]. The number of energy-specific open data sets has increased by much. On the other hand, the application domains, the electrical quantities measured, and the temporal resolution of available energy data show little variation across data sets. Most data sets describe the consumption of electrical energy for residential or office buildings, with a granularity of 10 or more minutes. For industrial production, which is responsible for a significant share of the overall energy consumption, open data is still scarce. Because of this, many challenges in this setting remain unattended.

To address this "data shortage", we present HIPE, a High-resolution Industrial Production Energy data set. HIPE contains smart meter readings of ten machines and the main terminal of an electronics production plant, over three months. The measurements are time series of various electrical quantities, e.g., active and reactive power, voltage, frequency, and harmonic distortion, with a resolution of 5 seconds. To our knowledge, this is the most comprehensive industrial energy data set openly available by now. It is accessible at https://www.energystatusdata.kit.edu/hipe.php.

In the second part of this article, we outline open challenges which become obvious when inspecting industrial energy data. This includes a discussion of characteristics of industrial machines and of differences to residential appliances. We describe potential use cases for energy data that are specific to these characteristics and where we expect HIPE to catalyze future research.

e-Energy '18, June 12-15, 2018, Karlsruhe, Germany

ACM ISBN 978-1-4503-5767-8/18/06…$15.00
https://doi.org/10.1145/3208903.3210278

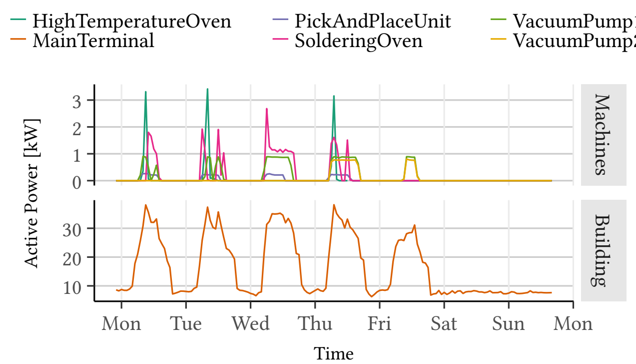

Figure 1: Week profile of active power consumption.

2 THE HIPE DATA SET

To make use of a data set, it is helpful to understand how one has collected the data and the meaning of the quantities measured. To introduce HIPE, we first describe the factory with its machines and production processes. We then present descriptive statistics to uncover important data characteristics.

2.1 Factory Setting

The Institute of Data Processing and Electronics (IPE) of Karlsruhe Institute of Technology (KIT) in Germany operates an electronics production site. It produces electronic systems for particle physics, battery systems, and medical applications in small batches, i.e., less than 1000 pieces. The production covers all processing steps from individual components to the final assembly.

Several machines and the factory building have been instrumented with high-resolution smart meters. The machines are either

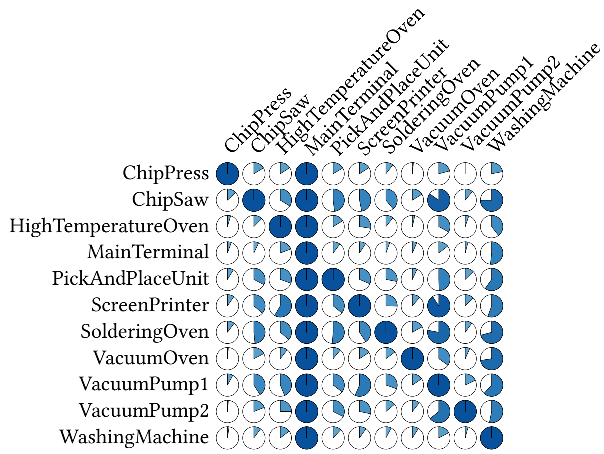

Figure 2: Simultaneous machine activity over thee months.

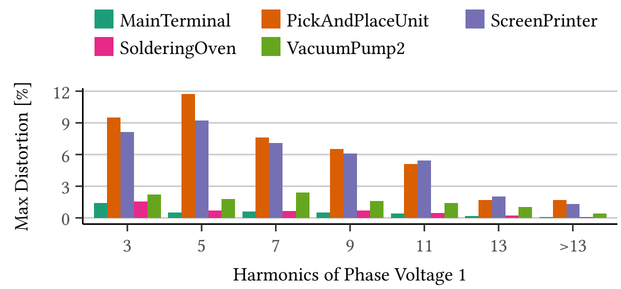

Figure 3: Maximum harmonic over three months.

connected to one (1P) or three phases (3P). They are located in a clean room (CR) or on the regular shop floor. They are as follows:

PickAndPlaceUnit (1P): Placement of electronic components, such as resistors and microcontrollers, on a printed circuit board (PCB). Energy consumption depends on the quantity of components per PCB and on the number of boards.

SolderingOven (3P): Components soldering to PCB. Energy consumption depends on throughput speed and temperature.

WashingMachine (3P): Cleaning of PCB. Energy consumption depends on temperature and process duration.

ScreenPrinter (1P, CR): Printing of material layers to interconnect electronic components via thick-film technology.

VacuumPump1 (3P) and **VacuumPump2 (1P):** Auxiliary machines to generate vacuum for other machines such as PickAndPlaceUnit. Energy consumption depends on vacuum demand.

HighTemperatureOven (3P, CR): Heats up to 1200 °C, fixing layers for thick-film technology. Energy consumption depends on temperature and heating duration.

VacuumOven (3P, CR): Oven with vacuum chamber.

ChipSaw (3P, CR): Separation of chips of a silicon wafer. Energy consumption depends on the wafer thickness.

ChipPress (3P, CR): Heat treatment of surfaces under high pressure, e.g., for multi-layered PCB. Energy consumption depends on pressure and temperature.

MainTerminal: Connection of factory to electrical grid. This also includes smaller machines, offices and air conditioning, which are not instrumented individually.

Production processes vary significantly between products. However, to illustrate a production cycle, we describe the production process of a data-acquisition module.

Example 2.1 (Production Cycle). In the first step, the PickAndPlaceUnit places electronic components on a PCB. During the placement, VacuumPump1 and VacuumPump2 supply negative pressure to manipulate components. Afterwards, the SolderingOven creates an electrical and mechanical connection between the components and the board. Lastly, the WashingMachine cleans the board and removes residue from previous productions steps.

2.2 Data Acquisition

All machines and the main terminal are monitored using EEM-MA600 [6] energy meters. A client polls the smart meters via MODBUS/TCP. After data is received, it proceeds with the next request. The poll time, and thus time between measurements, varies.

The smart meters measure over one hundred electrical quantities. A full data sheet is provided along with the data. Some quantities are not common with open energy-data sets, so we briefly describe their semantics and measurement specifics.

The machines are connected to an AC grid with three phases. The phases use the same *frequency* (F), which is the main indicator of grid stability. Deviations from the regulated frequency, 50 Hz in our case, indicate imbalance of demand and supply in the grid. The voltage between two phases is the *line voltage* (U), and the voltage between any phase and neutral is the *phase voltage* (V). The line voltages and phase voltages are regulated towards 230 V and 400 V respectively. However, loads inside and outside the factory can cause fluctuations. The *current* (I) of each phase is measured directly, and the current of the neutral conductor is inferred using Kirchhoff's Law. Since our measurements exceed the smart meter specifications of 5 A, we use a transformer[1] to scale down the measured currents.

In an AC system, the *active power* (P) of a phase is the load used by a machine to perform work. The *reactive power* (Q) is the power oscillating within the line. It is unusable for practical work. The *apparent power* (S) and *power factor* (L) are combinations of P and Q, to quantify the efficiency of the energy transfer in terms of the line capacity. Finally, *harmonics* are voltages and currents whose frequencies are multiples of the net frequency. The aggregation of harmonics in the grid is the *total harmonic distortion* (THD). Large THD can cause power quality issues.

2.3 Data Characteristics

To select suitable data analysis methods for a data set, one must understand its characteristics and distinctive features. We now describe main characteristics of HIPE and highlight some specifics.

General. HIPE contains three month worth of data, from October 2017 to December 2017. The temporal resolution is about 5 sec, with

[1]We use model PACT MCR-V1-21-44-75-5A-1.

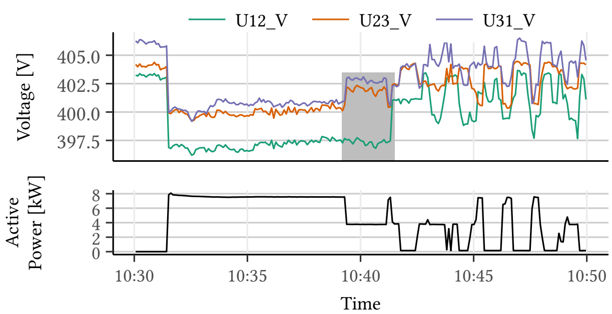

Figure 4: SolderingOven: Voltage and active power.

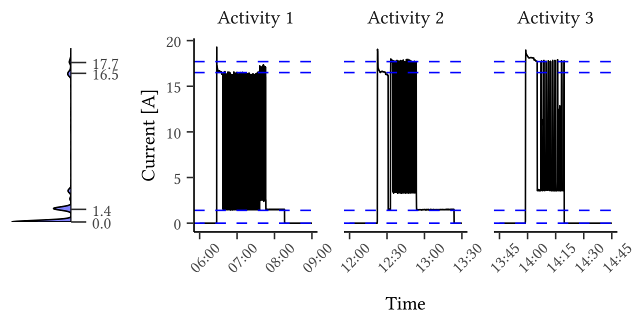

Figure 5: SolderingOven: Current during periods of activity.

a few exceptions for maintenance that increase the time between measurements to several minutes. This results in about 1.5 million measurements per machine and electrical quantity. In consequence, analysis methods for this data set must be robust to a non-constant time between measurements and should scale well with the number of measurements.

Factory Level. Summary statistics on the factory level give insights in production patterns. We illustrate this with three examples.

The first example is extracting consumption patterns and peak loads. Figure 1 depicts the power consumption over one week for the machines with the highest consumption as well as the main terminal. The maximum consumption is during working hours, and there is no production on weekends. The patterns and the peak loads are specific to the machine and vary between days.

The second example is analyzing machine run times. Figure 2 depicts simultaneous operation times of two machines, relative to the overall runtime of the machine in the row. Some machines often run simultaneously, which can be a result of process dependencies. For instance, VacuumPump2 (row) mostly is operating when VacuumPump1 (column) is active as well, but not vice-versa. This is plausible, because VacuumPump2 often supports VacuumPump1.

The third, more technical, example is analyzing harmonic distortions. Figure 3 depicts the maximum voltage harmonics measured over the three month period. In the past, PickAndPlaceUnit and ScreenPrinter have shown high distortions. This might be a good starting point to analyze the influence and propagation of harmonics among industrial machines and in a factory grid.

Machine Level. On a more fine-granular level, one can also study the behavior of individual machines. We illustrate this with examples from one day of the SolderingOven.

Figure 4 graphs the voltage and power consumption. During start-up, the voltage drops on all three phases. After some minutes, in the time interval highlighted in gray, the line voltages U23 and U31 increase, and U12 follows with a delay. This voltage and power pattern is typical for the SolderingOven. The machine can start up several times per day, depending on the production schedule.

Figure 5 depicts current measurements for three activity periods. The general shape of the current curve is similar for the different activity periods. However, the peak and the length of the period vary. There are four levels of current, at around 0 A, 1.4 A, 16.5 A, and 17.7 A, that have many sensor readings. These levels might characterize different machine states. For instance, the level at 1.4 A might indicate a stand-by level, which the machine enters at the end of Activities 1 and 2. So the high-resolution data allows to observe and analyze variation of machine behavior even within a day.

While the examples so far are simple, they allow to understand the data set and, consequently, factory operations and the local electrical grid. Future research can then come up with insights that indeed help operating and maintaining the production site.

3 RELATED WORK

Energy data sets have mainly been published for commercial and residential buildings. They differ by the temporal and spatial resolution as well as the electrical quantities collected. We now review energy-related data sets.

Commercial and Residential Buildings. The most prevalent data is power consumption of buildings, typically available in 10 min resolution. References distinguish between household and office buildings, because of differences in the consumption profile.

Numerous building data sets are publicly available [2]. REDD [13] is geared towards research on disaggregation of energy consumption. It provides 15 kHz current and voltage measurements and around 1 Hz power consumption of household appliances. Smart* [3] consists of several data sets, including consumption data, weather data and events such as motion sensor readings or HVAC switching, from three households, and power consumption of over 400 households with 1 min resolution.

The electrical data measured typically is restricted to load, current and voltage. Only one building data set features additional quantities such as line frequency and reactive power [15].

Industrial Data. Data sets collected from industrial settings are less numerous and mostly comprise consumption data with 1 to 60 min resolution. Only one industrial data set is publicly available [18]. It spans four months of 1 min resolution power consumption of a factory. Data is collected for four machines: a laser cutter, a bending machine, a robot welter and a laser shaper. A use case for this data is to find out the demand flexibility which the factory can offer to a virtual power plant.

However, there are other publications focusing on industrial settings for which the corresponding data is not publicly available. Examples are the generation of representative consumption profiles [12], finding instances of patterns identified by experts [9], clustering power-consumption curves in an industrial area [10] and non-intrusive load monitoring of heat pumps and compressors [11]. Hence, despite the "success" of data from buildings, data from industrial settings is scarce. This shortage is well-known, and several studies emphasize the need for open data, for reproducibility, and to develop new methods [7, 20].

4 FUTURE RESEARCH DIRECTIONS

The paradigm shift to an energy system where demand follows supply challenges industrial and residential energy consumers alike. However, many approaches to deal with this shift focus on residential data. They may not yield meaningful results when applied to industrial data. To strengthen this hypothesis, we first discuss characteristics of industrial energy data and highlight differences to residential sources. We then sketch use cases where we expect these characteristics to make a difference.

Device Inconsistency (C1). Machine programs vary for different products. Small batch sizes result in a frequent change of machine programs, which in turn causes varying load curves. This is different to the household settings, where many devices, e.g., a coffee machine, only serve one purpose and have stable signatures [8].

Instance Dissimilarity (C2). Different industry branches deploy different machinery. But even for the same branch, machine variant and wear can vary. Next, machines often are fully customized. So comparing two factory halls is unlikely to provide meaningful insights. Households in turn differ only in a few variables. With matching number of inhabitants and employment status, households tend to have similar consumption profiles. A reason is basic equipment like refrigerators or TV sets where load curves differ only slightly between model variants. For instance, most refrigerators switch on and off states to manage the temperature. The resulting load curve is similar for most models.

Inter-day Inconsistency (C3). Small-batch production induces frequent and irregular changes in production schedules, see Figure 1. For large-scale production, machine schedules might be more predictable, but still are subject to change. In contrast, residents have daily routines and often use appliances at similar times, e.g., the coffee machine in the morning. Changes to these routines are less common and might be subtle.

Device Dependence (C4). Production processes dictate the use of machines in specific sequences and impose time constraints between processing steps. This is different from households where appliances are mostly used independently from each other. The temporal correlation of appliances likely is an effect of daily routines, with exceptions like a washing machine followed by a dryer.

It is now interesting to probe how Characteristics C1-C4 affect methods that have been developed with residential data. Understanding these characteristics also might help to address industry-specific use cases. In the following, we discuss a selection of relevant use cases and indicate where HIPE might facilitate future research.

Non-Intrusive Load Monitoring. Non-intrusive Load Monitoring subsumes methods for the disaggregation of total power consumption to individual consumers. Methods developed on household data [4, 21] tend to rely on consistent device signatures and usage patterns. However, this assumption is unlikely to hold for industrial devices with varying signatures (C1) and flexible schedules (C3). Here, HIPE is a suitable benchmark because measurements are available for both individual appliances and the main terminal.

Load Forecasting. Consumption forecasts can improve the integration of renewable sources. Methods typically exploit regularities in consumption patterns [1, 19]. This is challenging for industrial consumption because of temporal and device inconsistencies (C1 and C3). HIPE spans three months of consumption data with changing production schedules and is therefore suitable to study these challenges for short-term and medium-term forecasting.

Load Simulation. Simulated loads are important to study characteristics of hypothetical scenarios, such as a residential district with high PV penetration. Current approaches use data generators to produce synthetic and at the same time realistic consumption profiles of individual households [16]. One assumption is that the appliance distribution and usage patterns are parameters of the generator and are known for typical households. In an industrial setting, it is not clear what constitutes a realistic consumption profile of a facility (C2). In addition, the simulation would have to consider process dependencies (C4). As HIPE provides both data from individual machines and a fully functional facility, it can serve as starting point for industrial load-simulation models.

Energy-Storage Integration. Energy-storage systems can reduce peak loads and offer services for frequency regulation [17]. To be used effectively, they have to provide appropriate reaction time, power and capacity. However, it is yet unclear what the exact requirements for industrial production are. To this end, HIPE can be helpful to specify these requirements and study how they depend on Characteristics C1 to C4.

Detecting Process Flexibility. Appropriate machine scheduling can increase renewable energy consumption and can reduce load peaks. Efficient scheduling relies on process flexibilities, but they often are not documented explicitly and are difficult to quantify. There has been only little research on learning process flexibility from energy data [14]. HIPE can serve as a foundation for respective research as it features strong process dependencies (C4) and frequent changes of production schedules (C3).

5 CONCLUSIONS

The availability of energy-related data sets has biased research towards residential sources. However, approaches developed on residential data may not produce meaningful results in an industrial setting. To encourage research on industrial energy data, we introduce HIPE, a comprehensive data set from an industrial production site. We expect HIPE to catalyze research on industrial applications and on the characteristics of energy data.

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG) as part of the Research Training Group GRK 2153: *Energy Status Data – Informatics Methods for its Collection, Analysis and Exploitation.*

REFERENCES

[1] Siddharth Arora and James W. Taylor. 2016. Forecasting electricity smart meter data using conditional kernel density estimation. *Omega* 59 (2016), 47–59. https://doi.org/10.1016/j.omega.2014.08.008

[2] Toktam Babaei, Hamid Abdi, Chee Peng Lim, and Saeid Nahavandi. 2015. A study and a directory of energy consumption data sets of buildings. *Energy and Buildings* 94 (2015), 91–99. https://doi.org/10.1016/j.enbuild.2015.02.043

[3] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. 2012. Smart*: An open data set and tools for enabling research in sustainable homes. In *SustKDD Workshop on Data Mining Applications in Sustainability*.

[4] Roberto Bonfigli, Stefano Squartini, Marco Fagiani, and Francesco Piazza. 2015. Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview. In *Environment and Electrical Engineering (EEEIC)*. 1175–1180. https://doi.org/10.1109/EEEIC.2015.7165334

[5] K. Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. 2013. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* 52 (2013), 213–234. https://doi.org/10.1016/j.enpol.2012.08.062

[6] PHOENIX CONTACT (Ed.). 2011. *User Manual EN EEM-MA600*. PHOENIX CONTACT, 586 Fulling Mill Road, Middletown, PA 17057, USA.

[7] Frank Eichinger, Daniel Pathmaperuma, Harald Vogt, and Emmanuel Müller. 2013. Data analysis challenges in the future energy domain. In *Computational Intelligent Data Analysis for Sustainable Development*. Chapman and Hall/CRC, 181–242.

[8] Linda Farinaccio and Radu Zmeureanu. 1999. Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses. *Energy and Buildings* 30, 3 (1999), 245–259. https://doi.org/10.1016/S0378-7788(99)00007-9

[9] Maël Guillemé, Laurence Rozé, Véronique Masson, Cérès Carton, René Quiniou, and Alexandre Termier. 2017. Improving Time-Series Rule Matching Performance for Detecting Energy Consumption Patterns. In *Data Analytics for Renewable Energy Integration: Informing the Generation and Distribution of Renewable Energy*. 59–71. https://doi.org/10.1007/978-3-319-71643-5_6

[10] Luis Hernández, Carlos Baladrón, Javier M. Aguiar, Belén Carro, and Antonio Sánchez-Esguevillas. 2012. Classification and Clustering of Electricity Demand Patterns in Industrial Parks. *Energies* 5, 12 (2012), 5215–5228. https://doi.org/10.3390/en5125215

[11] Emil Holmegaard and Mikkel Baun Kjaergaard. 2016. NILM in an Industrial Setting: A Load Characterization and Algorithm Evaluation. In *Smart Computing (SMARTCOMP)*. https://doi.org/10.1109/SMARTCOMP.2016.7501709

[12] José Antonio Jardini, Carlos M. V. Tahan, M. R. Gouvea, Se Un Ahn, and F. M. Figueiredo. 2000. Daily load profiles for residential, commercial and industrial low voltage consumers. *IEEE Transactions on Power Delivery* 15, 1 (2000), 375–380. https://doi.org/10.1109/61.847276

[13] J. Zico Kolter and Matthew J Johnson. 2011. REDD: A public data set for energy disaggregation research. In *SustKDD Workshop on Data Mining Applications in Sustainability*.

[14] Nicole Ludwig, Simon Waczowicz, Ralf Mikut, and Veit Hagenmeyer. 2017. Mining flexibility patterns in energy time series from industrial processes. In *Proceedings. 27. Workshop Computational Intelligence, Dortmund*.

[15] Stephen Makonin, Bradley Ellert, Ivan V. Bajić, and Fred Popowich. 2016. Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific Data* 3 (2016). https://doi.org/10.1038/sdata.2016.37

[16] Noah Pflugradt and B. Platzer. 2012. Behavior based load profile generator for domestic hot water and electricity use: Innostock. In *International Conference on Energy Storage (Innostock'12)*.

[17] Yuanyuan Shi, Bolun Xu, Di Wang, and Baosen Zhang. 2017. Using battery storage for peak shaving and frequency regulation: Joint optimization for superlinear gains. *IEEE Transactions on Power Systems* PP, 99 (2017). https://doi.org/10.1109/TPWRS.2017.2749512

[18] RWTH Aachen University. 2015. Smart Energy Data: Aachen/Cologne Smart Factory. (2015). https://data.lab.fiware.org/dataset/smart_energy_data-_aachen__cologne_smart_factory

[19] D. W. Van der Meer, Joakim Widén, and Joakim Munkhammar. 2018. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renewable and Sustainable Energy Reviews* 81, 1 (2018). https://doi.org/10.1016/j.rser.2017.05.212

[20] Kaile Zhou, Chao Fu, and Shanlin Yang. 2016. Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews* 56 (2016), 215–225. https://doi.org/10.1016/j.rser.2015.11.050

[21] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imran, and Sutharshan Rajasegarar. 2012. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors* 12, 12 (2012), 16838–16866. https://doi.org/10.3390/s121216838